Soil Physical Properties Web Database for GOSSYM and GLYCIM Crop Simulation Models

Ishtiaq Ali, Frank D. Whisler, Javed Iqbal,* Johnie N. Jenkins, and James M. Mckinion

ABSTRACT

Knowledge of soil physical properties is needed for various kinds of environmental studies, including crop simulation where the intended users are agronomists, consultants, and growers. However, acquiring and tabulating a complete set of soils analysis data for a particular region is expensive and laborious. This paper describes the construction of a web-based soil physical properties database to meet data requirements of users of cotton (*Gossypium hirsutum* L.) and soybean [*Glycine* max (L.) Merr.] crop simulation models (GOSSYM and GLYCIM, respectively) in addition to providing a generic data file of soil physical properties. The data are comprised of undisturbed samples of 1074 soil horizons (or about 300 sample sites) collected from farmers' fields using a tractor-mounted hydraulic probe. Standard laboratory analyses were performed to determine the various soil physical properties. An Oracle8 relational database management system was designed and implemented to store and deliver the soil physical properties using a client/server approach. A Perl database interface was used for the server-side database connectivity to build the soil files. The database accepts queries on several attributes, including sampling location, soil series name, state, county, farm, etc. This web-based database of soil physical properties was successfully implemented and tested and is available online at www.msstate.edu/~ia3/ (verified 11 Aug. 2004) without any restriction of identification or password.

Computer simulation models have been developed for a number of crops (Denisov, 2001; Whisler et al., 1986). For example, GOSSYM (*Gossypium* Simulator)/COMAX (Crop Management Expert) is a crop simulation model that estimates not only the growth and yield parameters of cotton, but also the levels of soil N, soil water, and crop stress factors as well as pertinent weather summaries (Hodges, 1992; McKinion, 1989; Landivar et al., 1983; Whisler et al., 1982). Output from COMAX includes recommendations for applications of irrigation, N, and plant growth regulators (Albers et al., 1992; McKinion et al., 2001; Baker, 1989). In soybean, the computer simulation model GLYCIM predicts crop growth and yield parameters (Acock et al., 1999; Haskett et al., 1995, 1997). Crop simulation models are increasingly being used outside the research community by cotton and soybean growers in the United States and in other countries including China, Greece, France, and Spain (Baker et al., 1983).

Most crop models rely on large amounts of data and information (Caldeira and Pinto, 1998), especially mechanistic types of models like GOSSYM and GLYCIM. In general, both dynamic and static types of data are required to run a crop model. A dynamic data set includes information about weather (minimum/maximum temperature, solar radiation, precipitation, and wind run), cultural practices (type and rate of fertilizer and herbicides; date and amount of irrigation), initial soil NO_3, percentage organic matter, and initial soil water content. A static data set includes information on crop variety and soil physical properties like soil bulk density, saturated hydraulic conductivity (K_s), soil moisture retention curve, and soil texture. Soil physical properties are generally treated as static properties unless fields are subjected to severe soil erosion, soil deposition, or aggressive land-leveling processes, and databases containing these data have been reported in the literature (Caldeira and Pinto, 1998; Tsuji et al., 1994; Carsel et al., 1991; Nemes et al., 2001). Schaap et al. (1998) calibrated a neural network using data sets of 1209 samples containing sand, silt, clay, and gravel contents; bulk density; and soil porosity to predict water retention parameters and values for K_s. Nemes et al. (2001) used Microsoft Access 97 to update and make user friendly the Unsaturated Soil Hydraulic Database (UNSODA) of some 790 soil samples from around the world that included data on water retention, hydraulic conductivity, and water diffusivity of the soils as well as pedological information. In a collaborative effort between 20 institutions in 12 European countries, Wösten et al. (1999) developed the Hydraulic Properties of European Soils (HYPRES), which is held within Oracle relational database management system (RDBMS). The HYPRES database contains both soil pedological and hydraulic data on 5521 soil horizons. Fitting the Mualem–Van Genuchten model parameters standardized the hydraulic data, and pedotransfer functions were used to predict the hydraulic properties from soil parameters collected during soil surveys.

Developing a database is an important first step in knowledge dissemination (Acock et al., 1997). The next step is making the database accessible to potential users with few constraints on final use of the data. This is possible today using computer hardware and software technologies and the World Wide Web. Pan et al. (2000) reported certain databases have been interfaced with crop growth models via the Internet, allowing the user to retrieve a database and then run a particular model over the Internet. In general, a database is organized

I. Ali and F.D. Whisler, Dep. of Plant and Soil Sci., Mississippi State Univ., Box 9555, Mississippi State, MS 39762; J. Iqbal, Dep. of Agric. and Biol. Eng., Mississippi State Univ., Mississippi State, MS 39762; and J.N. Jenkins and J.M. Mckinion, USDA-ARS Genet. and Precis. Agric. Unit, Mississippi State, MS 39762. Received 4 Feb. 2004. *Corresponding author (jiqbal@msa-msstate.ars.usda.gov).

Published in Agron. J. 96:1706–1710 (2004).

677 S. Segoe Rd., Madison, WI 53711 USA

Abbreviations: CGI, Common Gateway Interface; DBI, database interface; DBMS, database management system; GUI, graphical user interface; HTML, Hypertext Markup Language; K_s, saturated hydraulic conductivity; RDBMS, relational database management system.

and stored in a database management system (DBMS), consisting of a set of interrelated data and a collection of programs to retrieve the data (Codd, 1970). Compared with conventional file-based systems, the major advantages of a well-organized DBMS include prevention of data fragmentation and data dependence, support for queries and multiple data views, greater security, and support for concurrent access and data sharing. Several database models have been proposed, but the relational database model is most widely used (Harrington, 1988). The RDBMS is an implementation of the relational model that enables the user to define, create, control, and maintain a database.

Our objective was to design and implement a web-based database in a client/server system for the storage and retrieval of soil physical properties. A primary feature is a set of protocols for downloading "soil hydrology files" specific to either GOSSYM or GLYCIM models; however, the database also permits retrieval of "generic hydrology files" in ASCII format for use in other agricultural and environmental studies. All the data in the present study comprise actual values obtained from complete laboratory analysis of the various soil physical properties in each soil sample.

MATERIALS AND METHODS

Collection of Soil Samples

Since 1975, with initial development of GOSSYM and GLYCIM, the soil physics group with Mississippi State University has been collecting, analyzing, and recording soils data required to run these crop simulation models. Sampling was conducted mainly in the cotton-growing regions of the United States. Because GOSSYM and GLYCIM are mechanistic models, they require comprehensive and reliable information about several variables, including (i) crop variety, (ii) soil characteristics, and (iii) daily weather conditions. Information on crop variety and daily weather conditions are relatively easy to acquire whereas data sets about soil characteristics are harder to acquire and include important soil hydrologic properties, such as soil moisture retention curve, K_s, soil bulk density, and texture of each horizon to a 1-m depth (Khorsandi et al., 1997).

Most samples were collected from growers' fields where GOSSYM or GLYCIM test plots were already established. Once the major soils in the fields were identified, we obtained from three to five "replicate" cores from each soil series. Undisturbed soil profiles were taken to a depth of 1 m at each point using a Giddings hydraulic probe (Giddings Machine Co., Inc., Fort Collins, CO) mounted on the back of a tractor. Each 1-m profile was divided into three horizons, surface (A_p), subsurface (B_1), and deep (B_2), following soil pedological descriptions reported in soil survey (Soil Survey Staff, 1988) and used for soil particle analysis (Day, 1965). Using information on depth of each horizon, the Giddings probe was positioned within 1 m of the first sampling location and incrementally driven downward to obtain three undisturbed soil cores. Two samples were collected from each depth; one inside a 7.62-cm diam. by 1.00-cm ring was used for determination of the soil moisture release curve, based on seven pressure heads of 0.001, 0.01, 0.033, 0.067, 0.1, 0.5, and 1.5 MPa, and the second inside a 7.62-cm diam. by 7.62-cm ring was used for determination of K_s (Klute, 1986) and soil bulk density. The undisturbed soil cores from each horizon were placed in a tin can and sealed with electric tape to prevent moisture loss until ready for lab analysis.

System Design in a Client/Server Approach

The primary objective of the soil web database is to provide a user-friendly interface for the end user to download soil data files. In the soil web database development, the World Wide Web was used as an interface to soil database to retrieve and download soil data files. The Hypertext Markup Language (HTML), which is a language for creating web documents, was used to create web forms for the user's input request for the data files.

A plain HTML document that web servers provide is static, which means it exists in a constant state—a text file that doesn't change. To retrieve soil data on the web dynamically, the Common Gateway Interface (CGI) is used. The CGI is a method by which a web server can obtain data from (or send data to) databases, documents, and other programs and present that data to viewers via the web. A CGI program is executed in real time so that it can output dynamic information. Common Gateway Interface programs can be written in several programming and scripting languages, e.g., C/C++, PHP, TCL, Perl, etc. For the soil physical properties database, the Perl scripting language was used to write CGI scripts.

Database Design

The soil physical properties database was developed in the Oracle8 RDBMS, which we hosted on a UNIX operating system. A relational database is a collection of data items organized in a set of tables from which data can be accessed in different ways without rearranging or reorganizing database tables. This soil physical properties database consists of four tables titled as shown below in capital letters, with their corresponding column names in parentheses. Relationships represent the logical link between tables and are implemented in a relational database using primary and foreign keys. A primary key is a column or group of columns in a table that uniquely identifies records in that table. In relationships, the primary key of one table is linked with its equivalent, known as the foreign key, in a second table. For example, in this database, records in Tables A, B, and C can be selected by matching the value of the relevant foreign key in Table D with primary key values in the three tables.

A. LOCATION *(<u>location_id</u>, farm, county, state)*
B. TEXTURE *(<u>soil_tid</u>, tex_desc)*
C. SERIES *(<u>soil_sid</u>, series_desc)*
D. MODEL_FILE *(<u>location_fk</u>, <u>texture_fk</u>, <u>series_fk</u>, <u>year</u>, <u>Raw_file</u>, Gos_file, Gly_file)*

In this soil database, Tables A, B, and C function as lookup tables for the soil physical properties database. The underlined columns—*location_id* in LOCATION table, *soil_tid* in TEXTURE table, and *soil_sid* in SERIES table—hold unique numeric codes that are generated automatically and function as a primary key for their respective tables and as foreign keys in the MODEL_FILE table. The MODEL_FILE table has seven columns: *<u>location_fk</u>, <u>texture _fk</u>, <u>series_fk</u>, year, Raw_file*, *Gos_file*, and *Gly_file*. The attribute *Raw_file* contains the unique name of an archive file and contains up to five, but typically three, depth files generated from the laboratory analysis of each soil horizon. The *Raw_file* name also functions as the primary key for this table. The *Gos_file* column contains the names of the GOSSYM model file, and the *Gly_file* attribute will contain the names of the GLYCIM model files. Both *Gos_file* and *Gly_file* are derived from the raw file. The graphical user interface (GUI) client for the soil web database has a validation module, which takes care of all types of possible inconsistencies before uploading data to the server. Users can download all three types of files using a web browser.

Implementing Database on the World Wide Web

An interface was developed and implemented using the Perl database interface (DBI) module methods, which works as a bridge between the database and the user data request (Descartes and Bunce, 2000). The structured query language (SQL) is the DBI language used to access and manipulate data in the database. To retrieve soil physical properties records from the database, SQL statements, which are embedded in Perl DBI methods, are used. A user-friendly interface was developed in HTML, which will place a request to execute Perl CGI script.

a

b

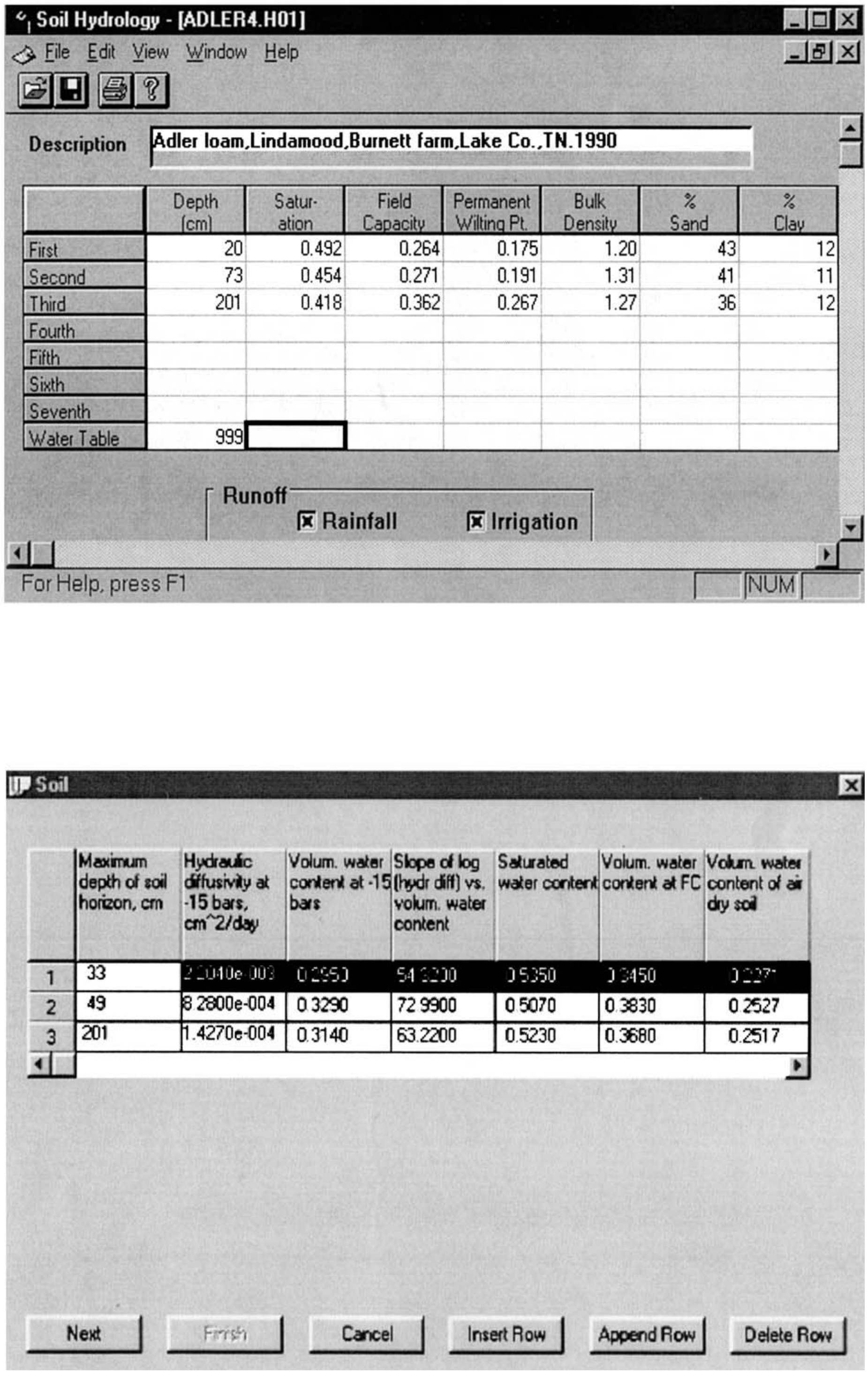

Fig. 1. Screen shot of downloadable (a) GOSSYM- and (b) GLYCIM-format soil hydrology files.

RESULTS

Database Features and Downloading Process

The website database is accessible with the URL http://www.msstate.edu/~ia3/ (verified 11 Aug. 2004) without any restriction of identification or password; however, only permitted administrator(s) can add, delete, or modify records in the database. A user can perform a search of the database using either the simple search or advanced search options. The simple search option allows one to search within different columns of state, county, soil series, and soil texture. The advanced search option allows the use of two additional fields, for instance, "sampling year" and "list order by." When "list order by" field is selected, the user is prompted to sort the results by an attribute, with the default sort order being by "sampling year." The advanced search combines the user's selected fields from the pull-down list with the Boolean operator "AND" to narrow the search.

Example

Suppose a soybean grower in north Mississippi is interested in leasing a field in Bolivar County, MS; a major soil in the field is mapped as Sharkey clay loam; and the grower is interested in downloading a file sampled in Bolivar County. The best approach to obtain the appropriate file from the database is to use the "Advanced Search" engine. After selecting "Advanced Search," the grower selects "Mississippi" from the STATE drop-down list, then "Bolivar" from COUNTY/STATE drop-down list, then "Sharkey" from SOIL SERIES drop-down list, and finally "Clay Loam" from SOIL TEXTURE drop-down list. Then the grower can select the desired file type from the different choices, GOSSYM, GLYCIM, and ASCII.

The GOSSYM hydrology file (Fig. 1a) has a line entitled "Description" for printing information on soil series name, sampling location, and sampling year. Soils data are arranged according to the number of horizons sampled and include the maximum depth of the horizon (cm); volumetric water content ($cm^3\ cm^{-3}$) at saturation, field capacity, and permanent wilting point; bulk density ($g\ cm^{-3}$); and percentage sand content and clay content. A typical GLYCIM hydrology file (Fig. 1b) is similar to that of GOSSYM with data arranged according to the different horizon depths sampled from the surface (cm). The variables for each horizon include hydrologic diffusivity at -1.5 MPa ($cm^2\ d^{-1}$), volumetric water content at -1.5 MPa, slope of log (hydraulic diffusivity) vs. volumetric water content, saturated water content, volumetric water content at field capacity, volumetric water content of air dry soil, bulk density ($g\ cm^{-3}$), the exponent in the water retention curve equation, K_s (cm d^{-1}), water potential of air entry (MPa), and percentage sand content and clay content. Both the GOSSYM/COMAX data file (*.hyd extension) and GLYCIM data file (*.soi extension) are in a format consistent with the respective crop model and include parameter estimates for any functional relationship between soil and water variables presently appropriate for each model. The user must have the model's soil editor installed and executed on their machines to upload and view the GOSSYM (*.hyd) and GLYCIM (*.soi) files.

Table 1. Format and description of raw data set.†

	Raw data	Description‡
1.	Commerce, silt loam,	soil series name, soil texture class
2.	Hood's Farm, Bolivar Co.,	sampling location
	MS, 1987	sampling year
3.	9.995, 0.980, –10, 0.37, 0.12	soil K_s, BD, pressure head, sand and clay content
4.	–333, 0.263, 24	pressure head, WC at field capacity and horizon depth (cm)
5.	0.137, 100, 0.0001	WC (air dry), number of iterations, allowable difference in WC at different pressure heads
6.	–1000, 0.195	pressure head, WC at the specified pressure head
7.	–10, 0.569	
8.	–100, 0.379	
9.	–333, 0.263	
10.	–667, 0.222	
11.	–1000, 0.195	
12.	–15000, 0.141	

† The table was first created in 70s for GOSSYM model from Lines 1 to 8, but later on with the development of GLYCIM model, Lines 9 to 12 were added; therefore, one will find duplicate data for volumetric water content at pressure head –333 on Lines 4 and 9 and for pressure head –1000 in Lines 6 and 11.

‡ K_s = saturated hydraulic conductivity (cm d^{-1}), BD = bulk density ($g\ cm^{-3}$), pressure head = centimeters of water column, and WC = volumetric water content ($cm^3\ cm^{-3}$).

Unlike the file formats used by the simulation models, the standard soil hydrology file is formatted as simple comma-separated, raw data (Table 1). Because raw output is in a compressed (zipped) file, the user must have a utility, such as WinZip, installed on their machines to uncompress and view the file contents. The raw file contains all measured soil parameters and is viewable in any DOS or WINDOWS text editor. A data entry GUI has been developed for additions or modifications to the data.

Testing and Evaluation

Construction of a client/server system involved several protocols, program languages, software tools, and platforms. When a sample data set was entered into each table, the functional constraints of the database were verified by applying basic transactions using Data Manipulation Language (DML) commands (e.g., select, insert, delete, and update) on the SQL command line interface. Next, a sample CGI script was developed to communicate with the soil database using the DBI module. The above transactions were verified by embedding DML commands into the CGI script. The results of both transactions by way of the SQL command line and by way of the DBI module were the same. Moreover, all transactions across the database returned the expected result.

REFERENCES

Acock, B., Y.A. Pachepsky, M.C. Acock, V.R. Reddy, and F.D. Whisler. 1997. Modeling soybean cultivar development rates, using field data from the Mississippi Valley. Agron. J. 89:994–1002.

Acock, B., Y.A. Pachepsky, E.V. Mironenko, F.D. Whisler, and V.R. Reddy. 1999. GUICS: A generic user interface for on-farm crop simulations. Agron. J. 91:657–665.

Albers, D.W., J. Mobley, and P.W. Tracy. 1992. Evaluation of GOSSYM-COMAX in Missouri. p. 1350–1352. *In* D.J. Herber and D.A. Richter (ed.) Proc. Beltwide Cotton Conf., Nashville, TN. 5–10 Jan. 1992. Natl. Cotton Counc. of Am., Memphis, TN.

Baker, D.N. 1989. Crop simulation for the analysis and management of cotton production systems: Overview and the status of the plant science research. p. 12–13. *In* J.A. Brown and D.A. Richter (ed.) Proc. Beltwide Cotton Prod. Conf., Nashville, TN. 2–7 Jan. 1989. Natl. Cotton Counc. of Am., Memphis, TN.

Baker, D.N., J.R. Lambert, and J.M. McKinion. 1983. GOSSYM: A simulator of cotton crop growth and yield. South Carolina Agric. Exp. Stn. Tech. Bull. 1089. Clemson Univ., Clemson, SC.

Caldeira, C., and P. Pinto. 1998. Linking DSSAT V3 to a relational database: The AGROSYS-DSSAT interface. Comput. Electron. Agric. 21:69–77.

Carsel, R.F., J.C. Imhoff, J.L. Kitlle, and P.R. Hummel. 1991. Development of a database and model parameter analysis system for agriculture soils. J. Environ. Qual. 20:642–647.

Codd, E.F. 1970. A relational model of data for large shared data banks. Commun. ACM 13(6):377–387.

Day, P.R. 1965. Particle fractionation and particle size analysis. p. 552–562. *In* C.A. Black (ed.) Methods of soil analysis. Part 1. 1st ed. Agron. Monogr. 9. ASA, Madison, WI.

Denisov, V.V. 2001. Development of the crop simulation system DIASPORA. Agron. J. 93:660–666.

Descartes, A., and T. Bunce. 2000. Programming the Perl DBI. 1st ed. O'Reilly Publ., Los Angeles, CA.

Harrington, J.L. 1988. Relational database management for microcomputers: Design and implementation. Holt, Rinehart, and Winston, Inc., New York.

Haskett, J.D., Y.A. Pachepsky, and B. Acock. 1995. Estimation of soybean yields at county and state levels using GLYCIM: A case study for Iowa. Agron. J. 87:926–931.

Haskett, J.D., Y.A. Pachepsky, and B. Acock. 1997. Increase of CO_2 and climate change effects on Iowa soybean yield, simulated using GLYCIM. Agron. J. 89:167–176.

Hodges, T. 1992. FORTRAN graphics library for PC-based crop simulation models. Agron. J. 84:124–128.

Khorsandi, F., M.Y.L. Boone, G. Weerakkody, and F.D. Whisler. 1997. Validation of the soil temperature subroutine HEAT in the cotton simulation model GOSSYM. Agron. J. 89:415–420.

Klute, A. 1986. Methods of soil analysis. Part I: Physical and mineralogical methods. Agron. Monogr. 9. ASA and SSSAJ, Madison, WI.

Landivar, J.A., D.N. Baker, and J.N. Jenkins. 1983. Application of GOSSYM to genetic feasibility studies: II. Analyses of increasing photosynthesis, specific leaf weight and longevity of leaves. Crop Sci. 23:504–510.

McKinion, J.M. 1989. Crop simulation for the analysis and management of cotton production systems: Expert systems. p. 13. *In* J.A. Brown and D.A. Richter (ed.) Proc. Beltwide Cotton Prod. Conf., Nashville, TN. 2–7 Jan. 1989. Natl. Cotton Counc. of Am., Memphis, TN.

McKinion, J.M., J.N. Jenkins, D. Akins, S.B. Turner, J.L. Willers, E. Jallas, and F.D. Whisler. 2001. Analysis of precision agriculture approach to cotton production. Comput. Electron. Agric. 32:213–228.

Nemes, A., M.G. Schaap, F.J. Leij, and J.H.M. Wösten. 2001. Description of the unsaturated soil hydraulic database UNSODA version 2.0. J. Hydrol. 251:151–162.

Pan, X., J.D. Hesketh, and M.G. Huck. 2000. OWSimu: An object-oriented and Web-based simulator for plant growth. Agric. Syst. 63:1933–1947.

Schaap, M.G., F.J. Leij, and M.T. Van Genuchten. 1998. Neural network analysis for hierarchical prediction of soil hydraulic properties. Soil Sci. Soc. Am. J. 62:847–855.

Soil Survey Staff. 1988. A basic system of soil classification for making and interpreting soil surveys. USDA-NRCS, Washington, DC.

Tsuji, G., G. Uehara, and S. Balas. 1994. DSSAT: A decision support system for agrotechnology transfer. Ver. 3. Vol. 2. Univ. of Hawaii, Honolulu.

Whisler, F.D., B. Acock, D.N. Baker, R.E. Fye, F.H. Hodges, J.R. Lambert, H.E. Lemmon, J.M. Mckinion, and V.R. Reddy. 1986. Crop simulation models in agronomic systems. Adv. Agron. 40:141–208.

Whisler, F.D., J.R. Lambert, and J.A. Landivar. 1982. Predicting tillage effects on cotton growth and yield. p. 179–198. *In* Predicting tillage effects on soil physical properties and processes. ASA, Madison, WI.

Wösten, J.H.M., A. Lilly, A. Nemes, and C. Le Bas. 1999. Development and use of database of hydraulic properties of European soils. Geoderma 90:169–185.